DOCUMENT RESUME

ED 353 289 TM 019 040

AUTHOR Hill, Richard
TITLE Assessments Developed in Support of Educational Reform in Contrast to Assessments Developed in Support of Educational Refinement.
PUB DATE Jun 92
NOTE 5p.; Paper presented at the Assessment Conference of the Education Commission of the States/Colorado Department of Educational Assessment (Boulder, CO, June 4, 1992).
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Accountability; *Educational Assessment; Educational Change; Educational Improvement; Elementary Secondary Education; Models; *Portfolios (Background Materials); State Programs; *Student Evaluation; Teacher Participation; *Test Construction; Testing Programs; Test Reliability; Test Validity
IDENTIFIERS Kentucky; *Reform Efforts

ABSTRACT

Educational refinement is defined as attempts to improve education through incremental improvements in the existing structure, while educational reform refers to change in the structure itself. A strong move is underway to reform education in the United States. It is argued that the development of assessments in support of educational reform requires assessment designs that will be as different as the designs for reform. Assessments developed in support of educational refinement have been highly concerned with issues of reliability. In the design of assessments developed in support of educational reform, the consequential validity of each item becomes an overriding consideration, and the impact on instruction is more important than technical precision. Because teachers and students will have to be daily participants in the process of educational reform, they will have to become involved in assessments. A model for an assessment system to support educational reform is proposed for Kentucky. In this design, portfolios are the primary means by which students are evaluated and schools held accountable. To ensure that portfolios reflect actual learning, the state can institute testing programs to determine which schools require on-site auditing of the portfolio process. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED353289

Assessments Developed in Support of Educational Reform in Contrast to Assessments Developed in Support of Educational Refinement

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

by
Richard Hill
Advanced Systems in Measurement and Evaluation, Inc.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
RICHARD K. HILL
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Presented at
1992 ECS/CDE Assessment Conference
June 4, 1992

Introduction--The Need to Reconceptualize Assessment

In reading this paper, it is important for the reader to distinguish between educational refinement and educational reform. The former activity is defined as one in which one is attempting to improve education through incremental inprovements in the existing structure; the latter as one in which one is trying to change the very structure itself. For the past several decades, education has been in a process of refinement. Assessments has been designed to support that activity. Currently, a strong movement is underway to reform education. Much thinking about new assessment programs have been extensions of existing programs, modified to be more consistent with educational reform, but still drawn from the basic orientation created by using assessments designed to assist educational refinement. It will be the argument in this paper that such thinking is confining, and has led to the design of assessments that are likely to be unmanageable and ineffective. The development of assessments in support of educational reform will require designs that are quite different; they will be designs that call for as much "mold breaking" in assessment as the designs for reformed educational programs are.

Differences between the Two Types of Assessment

Assessments developed in support of educational refinement examine the details of educational outcomes. Because they are looking for small, incremental improvements in test performance, they attend to issues of reliability carefully. They also tend to look at results of group performance, using statistics such as means and standard deviations, and sampling is often used to great advantage (either sampling of students or items). Assessments developed in support of educational reform must be quite different. Because the goals of educational reform typically are stated in terms of having all students perform to high levels of performance, these assessments must (1) define standards (high standards), (2) assess in a performance mode, and (3) assess and report on each student in terms of whether they have met the high standards.

Assessments developed in support of educational refinement are scalpels; assessment developed in support of educational reforms are sledgehammers. The former assessments are trying to uncover nuances of deficiencies in the existing system and provide teachers with the information to correct those deficiencies; the latter assessments are valuable to the extent that they are a factor in changing the entire system--i.e., support the reform.

TM019040

BEST COPY AVAILABLE

Assessments developed in support of educational refinement, quite appropriately, have been highly concerned with issues of reliability. Evidences of validity typically have been allowed to be relatively weak; that is, usual process has been to define the content domain, and argue that the test questions had been selected from across that domain. The model being used was that if the test met appropriate technical criteria--that is, if it was sufficiently scientifically defensible and therefore credible--the data from the assessment could be used to alter instruction, since those arguing for change would have the power of science to support them. In the design of an assessment developed in support of educational reform, the burden of proof becomes exactly the opposite. Each approach (or event within the assessment) must be justified in terms of its likely impact on instruction. Each event must be an illustration of the type of instruction that is to take place, and in fact, it can be effectively argued that the best questions <u>are</u> the best instruction (at least best as defined by advocates of the reform). Therefore, the consequential validity of each item becomes an overriding consideration. "The medium is the message." With each choice of question to be used in the assessment, we are defining for teachers unclear or uncertain about educational reform exactly what reformed education is to look like. Reliability, on the other hand, is not an issue of primary concern, as it is with assessment developed in support of educational refinement; it takes on a clearly secondary role. An assessment in support of educational reform only must be reliable to the extent that lack of reliability will negatively impact the affect of the assessment on instruction. That is, the main purpose of an assessment developed in support of educational reform is not to produce believable numbers that then may have an ensuing impact on instruction: it is to directly produce changes in educational practice. As a result, its design must be evaluated primarily in terms of its likely impact on instruction, not on its technical precision.

<u>Changes Needed in the Roles of Departments of Education and Teachers</u>

As noted above, it appears that just as an integral piece of educational reform is that it causes the roles of many participants in education to change (e.g., teachers change from information providers to managers), it will be crucial to reconceptualize the role of the major participants in the assessment. While some of the roles will remain the same (for example, the state department of education likely will determine the outcomes of education that will be valued, and establish and publish the standards of acceptable performance; teachers will be responsible for interpretation and daily application of those values), many roles must become markedly different.

For assessments developed in support of educational refinement, the state typically develops the exercises, administers them (usually with local support), collects the results, and then scores them. To the extent that performance exercises are included in the assessment, the state assumes responsibility for scoring them, either by having a contractor do it, or by selecting a group of teachers to accomplish the task. Because assessments developed in support of educational reform are trying to change the system by the involvement of teachers and students in the system, it is crucial that teachers and students become daily participants in the process. Therefore, in the design of such assessments, <u>all</u> teachers and students must participate in the creation of the problems that lead to the development of evaluatable products and <u>all</u> teachers must become involved in the scoring of those products (and maybe all students, as well--scoring their own work, the work of their peers, and perhaps the work of younger students).

Why Mere Modifications of Old Designs Cannot Work

It is important to note here, then, why assessments that are to support educational reform must be completely reconceptualized from earlier assessments. If one wanted to simply refine the old models of assessment to incorporate more performance oriented questions into them, it simply would not be practical to do so. Research has shown that performance events are greatly variable, and large numbers of them must be administered to obtain acceptably generalizable results. We estimate, for example, that a writing test needs to contain 6-10 prompts to generate reasonably reliable results. If each prompt take 45 minutes to administer and costs $2 to evaluate, a test for just that one content area would take 4½ to 7½ hours to administer and cost $12 to $20 just to score. These are expenses far beyond the budgets of current assessments--and most current leaders of reform in writing curriculum would argue that the test still does not carry the full message they would like. On-demand prompts administered in 45 minutes to not provide the opportunity to use the process they would like to see students employing. Thus, merely trying to force performance testing into the old models of assessment will not work; performance testing takes too much time to administer and costs too much to score to get sufficiently reliable data.

A Model Proposed for Kentucky

One model for an assessment system that will support educational reform is as follows. Under this system, portfolios become the primary means by which students are evaluated and schools are held accountable. In this paper, "portfolio" is defined to be a collection of evaluatable work aggregated over a period of time by students within constraints established by the state. The work in the portfolio should reflect students' habitual levels of performance, and is to be evaluated against concrete, common standards.

The assessment is practical since both the production and evaluation of students' work is merely a natural product of the reformed classroom. It also has great consequential validity, since it forces class time to be realigned to be consistent with reform efforts. For example, we received a letter from a teacher in Kentucky this spring in which the author provided what she felt was evidence that the effort to produce writing samples should be stopped. She cited such evidence as "Our students were at a disadvantage because they couldn't think of a topic since they never had to do that before," and "Some students got lower grades because we couldn't cover all the chapters in the textbook." Such arguments, of course, greatly reinforced our belief that requiring students to produce a writing portfolio was having great positive impact on the reform effort.

In the model we are proposing for Kentucky, teachers are integral to the system (and they must receive training reflective of the role they are to play). Along with their students, they determine, within constraints, what the assessment activities will be and have the responsibility of evaluating them. The role of the state, then, becomes that of an auditor. There must be at least two phases to the auditing: determining that the scores assigned to portfolios are accurate; and, determining that portfolios accurately reflect actual learning and are not an artifact of other events (e.g., parents writing the portfolios for their children).

Advanced Systems has developed systems for Kentucky and Vermont that reflect one means of accomplishing the first task--determining that each teacher has scored the portfolios accurately--practically. Since those models have been described elsewhere, detail will not be provided here. Key elements to them are that all teachers are audited, and the system provides feedback to all

teachers so that they can become sufficiently expert to articulate the standards and evaluate their students' products in a manner consistent with all other teachers.

To accomplish the second task, that of auditing to ensure that portfolios reflect actual learning, the state can institute a testing program that, on its surface, will look much like many current statewide testing programs. For Kentucky, for example, we have proposed a matrix sampled test that consists largely, if not exclusively, of open-ended questions. While this testing program may appear to be similar to those being developed in other states, the results will be used only to audit (that is, no one will be evaluated or held directly accountable for the results attained from this testing). Results for this on-demand testing will be used only to help determine schools that should receive auditing (we are expecting this will be involved on-site auditing). Because these tests differ in purpose from traditional state-operated, on-demand testing there are certain important consequences:

1. Consequential validity is not a direct concern of these tests. Since teachers and schools will not be held accountable for results on these tests, but rather their performance on the portfolios, it can be presumed that the message carried by these tests will not be nearly so strong as it would be if accountability decisions were to be made on the basis of them. Therefore, standards of authenticity can be somewhat relaxed (although the extent to which they can be relaxed is debatable).

2. Alternative methods of assessment then must be justified primarily in terms of their likelihood to lead to lower costs of auditing. Thus, for example, since more "authentic" means of assessment are likely to correlate higher with portfolios than multiple-choice items, they should be used if they can be accomplished at reasonable costs. Innovative efforts to assess achievement in language arts are being made by California and the New Standards Project. Such tests will be far too expensive to assess individual students, but they may prove to be cost effective if administered on a matrix sampling basis with an auditing purpose in mind. Again, these innovative assessments should be used not because of their influence on teaching and the reform thereof (since even the best of such tests will not be as "authentic" nor have the impact that involving teachers and students in portfolios will have), but because they are more likely to accurately identify situations where portfolios, even though scored accurately, are not likely to be reflective of actual learning.

3. Since auditing usually will be done at the class or school level (and not student by student), efficiencies such as matrix sampling can and should be built into the design of this auditing piece.